Dynamic generalized linear models with application
to environmental epidemiology

Monica Chiogna and Carlo Gaetan
Università di Padova, Italy

[Received February 2001. Revised May 2002]

Summary. We propose modelling short-term pollutant exposure effects on health by using dy-
namic generalized linear models. The time series of count data are modelled by a Poisson
distribution having mean driven by a latent Markov process; estimation is performed by the
extended Kalman filter and smoother. This modelling strategy allows us to take into account
possible overdispersion and time-varying effects of the covariates. These ideas are illustrated
by reanalysing data on the relationship between daily non-accidental deaths and air pollution
in the city of Birmingham, Alabama.

Keywords: Environmental epidemiology; Kalman filter; Randomized residuals; State space
models

1. Introduction

One of the issues at the centre of an on-going public and scientific debate concerns the question
of a direct and quantifiable short-term relationship between air pollution and the health of human
populations. Much of the evidence supporting conclusions in favour of a short-term associa-
tion between high pollution events and adverse health outcomes comes from environmental epi-
demiological studies employing time series regression of mortality and morbidity counts on a
variety of covariates including measures of pollution used as a proxy for the personal exposure of
individuals and a range of meteorological and confounding variables. Although some contro-
versial results have been highlighted among studies and among reanalyses of the same data
sets, results on the existence of statistical association between air pollution and health effects
seem to be fairly stable with respect to changes in modelling strategy and are accepted by the
epidemiological community.

The relevance of these investigations is witnessed by the effect that epidemiological evidence
has had in some countries in the establishment of national air quality standards. In fact, stan-
dards for particulate matter, i.e. materials of varying size and composition present in ambient
air in the form of solid particles or liquid droplets, have evolved in the last decade largely from
epidemiological studies (see, for example, US Federal Register (1997)).

Although epidemiologists have pointed to consistency across studies, only recently has the
statistical literature started to evaluate the effect of the methodology that is used in these studies
(see, for example, the special issue of Environmetrics, volume 11, part 6, 2000). Current opinion
is that many questions must be addressed before these studies can reasonably be said to assess
the association between air pollution and health and therefore to justify their use in determining

Address for correspondence: Monica Chiogna, Dipartimento di Scienze Statistiche, Università di Padova, Via
Cesare Battisti 241, 35121 Padova, Italy.
E-mail: monica@stat.unipd.it

© 2002 Royal Statistical Society 0035–9254/02/51453
public policy. Controversial issues include the treatment of serial dependence of the data, the in-
fluence on the estimated associations of long-term trends, possible confounding effects between
air pollutants and weather, and the possible existence of a threshold level for air pollutants,
below which there is no observable effect. For these and other reasons, the construction of
statistical models is particularly difficult.

In this paper, we focus in particular on handling the dynamic features of the data. In epidemi-
ological time series, serial dependence is due to different sources. Firstly, explanatory variables
are autocorrelated, so the time series structure of the covariates imparts a highly structured
pattern of interdependence on the response. Then, the effect of the explanatory variables on the
outcomes usually lasts for some time; for example, the effect of a high pollution event on the
health of a population carries over for a few days, although the dose–response mechanism is un-
fortunately unknown. This, again, depends partly on the autocorrelation of the pollutants, and
partly on the biologically plausible response of living organisms to exposure to air pollutants
(Clarke et al., 1999).

A natural way to deal with such issues would be to develop association models in which the
dependence structure within the explanatory variables and between covariates and the response
is correctly accounted for. However, in most cases neither the dependence mechanism on the
explanatory variables nor adequate knowledge of the physical association between the response
and covariates is known, so an explicit probabilistic model of association is not available. There-
fore, the accepted statistical strategy for the construction of the model consists of three stages:
first, the control of long-term trend and seasonal variations is performed by parametric filtering
or nonparametric smoothing. Then, confounding variables (such as meteorology) are included
in the predictor and the best set of confounders is selected by making use of statistical or ad
hoc criteria. Finally, pollutants are added to the model in different ways to account for the
‘carry-over’ effect.

Our paper is about the use of an alternative strategy for dealing with all the temporal fea-
tures of the series, which captures long-term, seasonal cycles and delayed effects. Our proposal
is to tackle the complex modelling task by making use of dynamic generalized linear mod-
els (DGLMs; Fahrmeir and Tutz (1994)). We consider regression models based on state space
models for the outcomes; serial dependence is added to the model structure by making use of
random coefficients supplemented by processes, such as random walks, that can handle the
autocorrelation. To our knowledge, DGLMs have never been applied in the context of short-
term health effect studies before, although some dynamic models for mortality have already
been developed in different contexts. Manton (1991) and Manton et al. (1994) have developed
a suite of models that relate mortality explicitly to individual risk factors (such as physiological
or behavioural characteristics). Manton’s models, however, do not fit the observational nature
of the environmental applications.

We contend that the dynamic approach described herein provides a flexible methodology
which more neatly accounts for the dependence that is brought into the models by the temporal
behaviour of the covariates, and which can be tailored to a variety of situations. From a practical
viewpoint, we do not claim that the dynamic approach is superior to standard methodology in
giving more reliable estimates of effects; nor do we aim to solve any of the controversy about
the interpretation of evidence. Instead, we would like first to highlight how all strands related
to the temporal features of the series can be brought together into a coherent framework and
second to show that modelling strategies that are different from those adopted in current prac-
tice are worth considering and are informative. For these purposes, we reanalyse data on the
relationship between daily non-accidental deaths and air pollution in the city of Birmingham
(Alabama) in the period 1985–1988. This data set was first treated by Schwartz (1993) and has
been extensively reanalysed since then (Smith et al., 2000), giving rise to a lively debate about
the link between particulate matter and mortality.

The remainder of this paper is organized as follows. In Section 2, we highlight some weaknesses of standard approaches which lead us to consider DGLMs. In Section 3, we briefly describe the Birmingham case-study. In Section 4, DGLMs are briefly reviewed. Moreover, the definition of randomized residuals, which are used for model checking, is recalled. In Sections 5 and 6 respectively we describe the application of a DGLM to our data and discuss the consequent results. In Section 7, we discuss the advantages and disadvantages of the modelling strategy.

2. Why move to dynamic models?

In environmental epidemiology, the usual modelling strategy is based on estimating proper generalized linear models (GLMs; McCullagh and Nelder (1989)) or generalized additive models (Hastie and Tibshirani, 1990) with the assumption of independent outcomes. Let \( Y_t; t = 1, \ldots, T \), be observations from an exponential family with mean \( \mu_t \) where \( x_t = (x_{t1}, \ldots, x_{tp})' \) is a \( p \)-dimensional vector of explanatory variables. The usual GLM framework assumes that the mean \( \mu_t = \mu(x_t) \) relates to a linear predictor via an invertible link function:

\[
\mu_t = h \left\{ \sum_{m=1}^{p} \beta_m \phi_m(x_{mt}) \right\},
\]

where the \( \phi_m \) are known transformations of the explanatory variables and the coefficients \( \beta_m \) are the unknown quantities of interest. In the generalized additive model context, the coefficients \( \beta_m \) are considered to be fixed equal to 1 and the \( \phi_m \) are unknown smooth functions to be estimated.

After the models have been constructed following the stages described in Section 1, a check on the residuals’ autocorrelation is performed to detect violations of the independence assumption. It is generally thought that, after removing long-term trends and seasonal patterns and adjusting for other covariates such as temperature, the autocorrelation of residuals should typically be weak and therefore acceptable. However, if in the final model this autocorrelation is substantial, then ad hoc corrections are introduced to account for it.

In the literature, a variety of approaches have been proposed to model autocorrelated data in the Gaussian context, but extensions to non-normal outcomes is not trivial. Two main approaches can be followed to add autocorrelation to the standard GLM or generalized additive model setting: either a latent autocorrelated time series error is assumed for the model (parameter-driven model), which means that correlation between two subsequent outcomes is a known function of the marginal means of the outcomes and perhaps of some additional parameters, or correlation is inserted into the model by making the current outcome depend explicitly on past outcomes (observation-driven model). Recently, Brumback et al. (2000) discussed how transitional regression models, meaning non-linear regression models expressed in terms of conditional means and variances given past observations, provide a unifying framework for the two mainstream approaches to extending the GLM for autocorrelated data.

In the environmental literature, various studies have appeared which make use of the above-mentioned approaches to account for autocorrelation (see for example Brumback et al. (2000)). Despite the numerous successful applications, the mechanism on which this strategy relies to account for autocorrelation seems unsatisfying. To see why, consider a non-negative process \( \{\varepsilon_t\} \) such that

\[
Y_t|\varepsilon_t, x_t \sim \text{Po}(\varepsilon_t \mu_t)
\]
where \( \text{Po}(\mu) \) denotes a Poisson distribution with mean \( \mu, \mu_t = \exp(\mathbf{x}_t'\beta) \), and \( \beta' = (\beta_1, \ldots, \beta_p) \) is a vector of regression coefficients. It is further assumed that \( \{\varepsilon_t\} \) is a stationary process with mean \( E(\varepsilon_t) = 1 \), variance \( V(\varepsilon_t) = \sigma^2 \), autocovariance function \( \gamma_\varepsilon(h) = E\{(\varepsilon_t - 1)(\varepsilon_{t+h} - 1)\} \) and autocorrelation function \( \rho_\varepsilon(h) = \gamma_\varepsilon(h)/\sigma^2 \). This specification of a latent process acting on the conditional mean of the Poisson counts is suggested in Zeger (1988) and used in Campbell (1994). Now assume that \( \varepsilon_t = \exp(\alpha_t) \). Under the assumption of a stationary Gaussian latent process \( \{\alpha_t\} \), it is easy to show the following facts about the first and second moment of the observed process \( Y_t \):

\[
E(Y_t) = \mu_t, \\
V(Y_t) = \mu_t + \sigma^2\mu_t^2, \\
\text{corr}(Y_t, Y_{t+h}) = \frac{\rho_\varepsilon(h)}{\sqrt{\{1 + (\sigma^2\mu_t)^{-1}\}\{1 + (\sigma^2\mu_{t+h})^{-1}\}}}.
\]

Therefore,

\[
0 < |\text{corr}(Y_t, Y_{t+h})| < |\text{corr}(\varepsilon_t, \varepsilon_{t+h})|,
\]

which implies that the autocorrelation function for the Poisson count process is dominated by that for the underlying latent process. This last observation illustrates the difficulty in detecting serial dependence by using standard residuals: little or no correlation in the \( Y_t \), and therefore in the residuals, may mask significant correlation in the latent process, which demonstrates the weakness of developing identification procedures based solely on second-order properties of the data.


The relationship between mortality and air pollution for the city of Birmingham, Alabama, in the period from January 1st, 1985, to December 31st, 1988, was first explored by Schwartz (1993), who found a statistically significant effect of PM10 on mortality, where PM10 is defined as particulate matter with a mass median aerodynamic diameter less than 10 \( \mu \text{m} \). The study was then repeated by Samet et al. (1995), who developed new computational algorithms for fitting the models to the same data set and essentially confirmed Schwartz’s results. On a different data set for a different period, Roth and Li (1996) fitted a wide variety of models, claiming that there was no evidence of any association between particulate matter and mortality in Birmingham.

Recently, Smith et al. (2000) built a new data set for the period from August 3rd, 1985, to December 31st, 1988, based on daily mortality data, PM10 and various meteorological data. After a careful analysis, they argued

‘that the results may depend on a number of ad hoc features of the analysis, including which meteorological variables to adjust for, and the manner in which different lagged values of particulate matter are combined into a single exposure measure’.

Our analysis is based on the data set of Smith et al. (2000); we refer to Smith et al. (2000) for a detailed description of the variables included in the data set and on their construction. Daily mortality data count the daily non-accidental deaths for Birmingham residents aged 65 years and over. Table 1 includes summary statistics for the daily air pollutant concentration, key meteorology and death counts recorded in Birmingham during the study period.

Fig. 1 shows the time series of daily counts for the study period. For exploratory purposes,
Table 1. Summary statistics for daily air pollutant concentration and deaths counts, Birmingham, Alabama, August 3rd, 1985–December 31st, 1988

<table>
<thead>
<tr>
<th>Statistic</th>
<th>PM10 ($\mu g m^{-3}$)</th>
<th>Tmax ($^\circ C$)</th>
<th>Tmin ($^\circ C$)</th>
<th>Humidity (g kg$^{-1}$)</th>
<th>Mortality (counts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>47.23</td>
<td>23.39</td>
<td>10.65</td>
<td>9.31</td>
<td>15.06</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>23.77</td>
<td>8.61</td>
<td>8.91</td>
<td>4.89</td>
<td>4.25</td>
</tr>
<tr>
<td>Minimum</td>
<td>8.00</td>
<td>$-3.30$</td>
<td>$-12.20$</td>
<td>0.80</td>
<td>3.00</td>
</tr>
<tr>
<td>Lower quartile</td>
<td>29.00</td>
<td>17.20</td>
<td>3.30</td>
<td>4.90</td>
<td>12.00</td>
</tr>
<tr>
<td>Median</td>
<td>44.00</td>
<td>24.40</td>
<td>11.10</td>
<td>8.70</td>
<td>15.00</td>
</tr>
<tr>
<td>Upper quartile</td>
<td>59.30</td>
<td>31.10</td>
<td>19.40</td>
<td>14.00</td>
<td>18.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>163.00</td>
<td>38.30</td>
<td>25.60</td>
<td>18.30</td>
<td>32.00</td>
</tr>
</tbody>
</table>

Fig. 1. Time series of the observed daily death counts along with the LOESS smoother of the counts (-----) and of the PM10 values (- - - - -)

the LOESS smoothed series (Cleveland, 1979) of PM10 concentrations has been added to the plot. It appears that there is seasonality in the mortality series, although the periodicity is not regular and there is not a clear correlation between the behaviour of PM10 and mortality.

To develop our models, we shall use the above-mentioned study in constructing and selecting covariates, although we do not aim at building models which are comparable with the models that have already been published. In fact, one of our aims is to explore the extent to which the DGLM setting allows the construction of models which are ‘more parsimonious’ than the traditional models, where parsimony is defined in terms of the number of explanatory and confounding variables that need to be included in the model and not with respect to the number of parameters to be estimated.
4. Dynamic generalized linear models

4.1. Model definition

Consider a series of counts \( \{Y_t\} \) recorded at equally spaced times \( t = 1, \ldots, T \) along with a vector of covariates \( \mathbf{x}_t \in \mathbb{R}^p \). Assume that it is reasonable to divide the vector of covariates into two components \( \mathbf{x}_t = (\mathbf{x}_{1t}^t, \mathbf{x}_{2t}^t)' \), where the first component \( \mathbf{x}_{1t} \) includes covariates which we expect to contribute to the intrinsic mean tendency of the counts and the second component \( \mathbf{x}_{2t} \) includes perturbing factors, whose influence can be thought of as being constant over time. If the central tendency of the counts may be thought of as resulting from the influence of both types of covariate, then it is reasonable to model the effect of \( \mathbf{x}_{1t} \) by a univariate random process \( \phi_t = \phi_t(\mathbf{x}_{1t}) \) and to treat the effects of \( \mathbf{x}_{2t} \) as fixed effects. Therefore, we assume that the conditional distribution of \( Y_t \) given \( \phi_t \) follows the Poisson distribution

\[
Y_t | \phi_t \sim \text{Po}(\exp(\phi_t + \mathbf{x}_{2t}' \beta_2))
\]

with \( \beta_2 \) representing the unknown regression coefficients for \( \mathbf{x}_{2t} \).

The random process can take on different specifications. In the simplest case, it includes only a random trend. Simple random trend models are for example random walks of first order, \( \phi_t = \omega_t, \omega_t = \omega_{t-1} + \delta_t \), where \( \delta_t \) is an independent Gaussian random variable, i.e. \( \delta_t \sim N(\mu, \sigma^2) \), or of second order, \( \phi_t = \omega_t, \omega_t = 2\omega_{t-1} - \omega_{t-2} + \delta_t \). According to the second formulation, \( \phi_t \) is the discrete version of a spline for a trend component, so model (1) resembles a semiparametric model.

In more complex situations, the random process includes long-term covariates as well. An example with a covariate \( x_{1t} \) is

\[
\phi_t = \omega_t + x_{1t} b_t, \\
\omega_t = 2\omega_{t-1} - \omega_{t-2} + \delta_t, \\
b_t = b_{t-1} + \xi_t
\]

where \( \xi_t \sim N(\mu, \sigma^2) \) is a Gaussian random variable uncorrelated with \( \delta_t \). In this second formulation, the coefficient of \( x_{1t} \) varies dynamically in time. This is a crucial aspect of the model. Firstly, it is possible to express long-term changes in the effects of the covariates. Secondly, the dynamic on the coefficients captures the carry-over effect of the covariates which usually causes the effect at time \( t \) to be influenced by covariates at previous times. This seems highly relevant to the applications that we consider.

The previous setting falls into the general framework of DGLMs. In general terms, DGLMs relate the series \( \{Y_t\}, t = 1, \ldots, T \), to a sequence of unobservable states or parameters \( \alpha_t \in \mathbb{R}^q, t = 0, \ldots, T \). The relationship is given by an observation model

\[
Y_t | \alpha_t \sim p(y_t | \alpha_t) = c_t(y_t) \exp\{\theta_t y_t - b_t(\theta_t)\}
\]

where \( \theta_t = g(\eta_t) \), the natural parameter, is a function of \( \eta_t = \mathbf{z}'_t \alpha_t, c_t(\cdot) \) and \( b_t(\cdot) \) are known functions and \( \mathbf{z}_t \) is an \( n \times 1 \) design vector.

Similarly to static GLMs, the conditional mean \( E(Y_t | \alpha_t) = \mu_t(\alpha_t) \) is linked to the linear predictor \( \mathbf{z}_t' \alpha_t \) by

\[
\mu_t(\alpha_t) = h(\eta_t),
\]

where \( h(\cdot) \) is an appropriate link function. In our applications, where log-linear Poisson models are natural choices, the observation model is

\[
p(y_t | \alpha_t) = \exp\{\theta_t y_t - \exp(\theta_t)\}/y_t!,
\]

where \( \theta_t \) is thought of as being constant over time.
with \( \theta_t = z'_t \alpha_t \) and \( h(z'_t \alpha_t) = \exp(z'_t \alpha_t) \).

The observation model is supplemented by a Gaussian transition model

\[
\alpha_t | \alpha_{t-1} \sim N(F_t \alpha_{t-1}, Q_t),
\]

(4)

with transition matrix \( F_t \) and initial state \( \alpha_0 \sim N(\mathbf{a}_0, Q_0) \), where \( \mathbf{a}_0, Q_0, Q_t \) and \( F_t \) are hyperparameters. This implies that the sequence \( \{ \alpha_t \} \) forms a Gaussian Markov process that does not even need to be stationary.

Setting (4) gives rise to a very broad class of transition models. For example, consider model (1) with structure defined in equation (2) and assume, for convenience, that \( x_{2t} = x_{2t} \).

Then,

\[
\alpha_t = \begin{pmatrix}
\omega_t \\
\omega_{t-1} \\
b_t \\
\beta_2
\end{pmatrix} = \begin{pmatrix}
2 & -1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
\omega_{t-1} \\
\omega_{t-2} \\
b_{t-1} \\
\beta_2
\end{pmatrix} + \begin{pmatrix}
\delta_t \\
0 \\
0 \\
0
\end{pmatrix} = F_t \alpha_{t-1} + \mathbf{u}_t,
\]

where \( \mathbf{u}_t \sim N(\mathbf{\mu}_t, \text{diag}(\sigma_{\xi_1}^2, 0, \sigma_{\xi_2}^2, 0)) \) and \( \beta_2 \) is the regression coefficient for \( x_{2t} \). To obtain model (2), we set \( \mathbf{z}_t = (1, 0, x_{1t}, x_{2t})' \).

### 4.2. Interpretation of the hyperparameters in a standard regression setting

It is desirable to choose the hyperparameters of the latent process so that the marginal mean of the response variable enables an interpretation that is similar to that in standard settings. For example, in the Poisson case we would like to have

\[
\mu_t(t) = \exp(x_t' \beta).
\]

In this way, parameters in the marginal mean retain the usual interpretation and are directly comparable with estimates obtained through traditional approaches.

To exploit marginal comparability, consider again model (2). On setting \( b_0 = \beta_1 \) and \( \xi_t \sim N(-\sigma_{\xi_2}^2 x_{1t}/2, \sigma_{\xi_2}^2) \), \( \omega_0 = \omega_1 = \beta_0 \) and \( \xi_t \sim N(-\sigma_{\xi_2}^2 (2t - 1)/6, \sigma_{\xi_2}^2) \), it can be shown that

\[
b_t \sim N\left(\beta_1 - \frac{t}{2} \sigma_{\xi_2}^2 x_{1t}, 1 \sigma_{\xi_2}^2\right)
\]

and

\[
E(Y_t) = \exp(\beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t}) = \exp(x_t' \beta).
\]

In this model, \( \beta_1 \) plays the role of the overall mean effect of \( x_{1t} \) in the study period and retains the interpretation of relative risk per unit change in the covariate as in standard GLM settings. Note that the variance of \( b_t \) depends on \( t \), which suggests that small values should be set for \( \sigma_{\xi_2}^2 \) to control the magnitude of the variance of \( b_t \).

### 4.3. Estimation procedure

DGLMs have two unknown quantities: the unobservable states \( \alpha_t \) and the hyperparameters \( \mathbf{a}_0, Q_0, Q_t \) and \( F_t \). We summarize the hyperparameters in the vector \( \lambda \) and we assume for the moment that \( \lambda \) is fixed and known. Let \( \alpha^*_t = (\alpha'_0, \ldots, \alpha'_t)' \) and \( y^*_t = (y_1, \ldots, y_t)' \), and let the conditional distribution of \( \alpha^*_T \) given \( y^*_T \), be

\[
p(\alpha^*_T | y^*_T) = \prod_{t=1}^T p(y_t | \alpha_t) \prod_{t=1}^T p(\alpha_t | \alpha_{t-1}) p(\alpha_0)/p(y^*_T).
\]
As an estimate for $\alpha_T^*$, we consider the conditional mode

$$\hat{\alpha}_T^* = \arg \max_{\alpha_T^*} \{ p(\alpha_T^*|y_T^*) \}. $$

Because of the complicated form of the conditional distribution, inference requires some approximations. Although many Markov chain Monte Carlo algorithms have been proposed for dealing with this problem (Shephard and Pitt, 1997), we have chosen the approach proposed by Fahrmeir and co-workers (Fahrmeir and Kaufmann, 1991; Fahrmeir, 1992; Fahrmeir and Wagenpfeil, 1997). As $p(y_T^*)$ does not depend on $\alpha_T^*$, maximizing $p(\alpha_T^*|y_T^*)$ is equivalent to maximizing the function

$$PL(\alpha_T^*) \equiv \sum_{t=1}^{T} \log \{ p(y_t|\alpha_t) \} - \frac{1}{2} \sum_{t=1}^{T} (\alpha_t - F_t \alpha_{t-1})' Q_t^{-1} (\alpha_t - F_t \alpha_{t-1})$$

$$- \frac{1}{2} (\alpha_0 - a_0)' Q_0^{-1} (\alpha_0 - a_0).$$

Maximization of function (5) is accomplished by the working Kalman filter and smoother (WKFS) algorithm (see Appendix A for details).

An interesting interpretation is to consider expression (5) as a penalized log-likelihood function where $\alpha_t$ can be considered as a fixed but unknown parameter. From this viewpoint, the first term in function (5) measures the goodness of fit obtained by the linear predictor $z_t' \alpha_t$, whereas the remaining terms penalize the roughness of the fit. This suggests that estimation of the unknown hyperparameters $\lambda$ can be accomplished by employing a cross-validation criterion (Fahrmeir and Wagenpfeil, 1997) (see Appendix A for details).

4.4. Diagnostics

The computation of residuals deserves some care. In principle, we could use Pearson-type residuals:

$$p_t = \frac{Y_t - \mu_t(\hat{\alpha}_t)}{\sqrt{\Sigma_t(\hat{\alpha}_t)}}. $$

Unfortunately, because of the dynamic set-up, these residuals do not have the usual properties. More appropriate inference is accomplished by considering the one-step-ahead predictive distribution of a future value $Y_t$, given observations up to $t - 1$,

$$p_t(s) = \Pr(Y_t \leq s|y_{t-1}^*).$$

Following Smith (1985), we define the following randomized residuals for time series of count data:

$$v_t = \Phi^{-1}(r_t)$$

where $r_t = (1 - u_t) p_t(y_t - 1) + u_t p_t(y_t)$, $u_t$ is a uniform random variable and $\Phi$ is the standard normal distribution. $r_t$ will be called the $P$-score and $v_t$ will be called the transformed $P$-score. Distribution (6) is known analytically only when the observation model is Gaussian. However, we can represent expression (6) as an infinite mixture distribution:

$$p_t(y_t) = \int_{-\infty}^{\infty} \Pr(Y_t \leq y_t|\theta) p(\theta|y_{t-1}^*) \, d\theta.$$
which can be approximated by a finite mixture distribution (Frühwirth-Schnatter, 1996). We substitute the exact density \( p(\theta_t | y_{t-1}) \) by a normal density with first two moments \( \theta_{t|t-1} = z_t \alpha_{t|t-1} \) and \( \Theta_{t|t-1} = z_t V_{t|t-1} z_t' \) and then we use Gauss–Hermite integration

\[
p_t(y_t) \approx \frac{1}{\sqrt{\pi}} \sum_{i=1}^{M} \Pr(Y_t \lesssim y_t | \theta_{t}^{(i)}) \omega_{M}^{(i)}
\]

with

\[
\theta_{t}^{(i)} = \theta_{t|t-1} + (2\Theta_{t|t-1})^{(i)} \tau_{M}^{(i)}
\]

where \( \omega_{M}^{(i)} \) and \( \tau_{M}^{(i)} \) are the weights of univariate Gauss–Hermite integration of order \( M \) (Press et al., 1992). After the computation of the transformed \( P \)-scores, standard diagnostic tools can be used. A plot of the transformed \( P \)-scores against time can be used to detect outliers and heterogeneity. A normal probability plot of transformed \( P \)-scores can be useful as well for detecting model inadequacies. A check for autocorrelation can be performed by exploring the empirical autocorrelation function.

5. Modelling in practice

Our model building strategy starts from the ‘full’ model, i.e. the model including all the key covariates, and attributing time-varying coefficients to all the covariates. Then, model selection proceeds by removing non-significant variables, i.e. variables whose corresponding time-varying confidence bands always include 0, and by treating the effects of covariates, whose estimated confidence bands do not show evidence of time-varying behaviour, as fixed, i.e. non-random, effects.

We describe the modelling procedure in detail. As mortality data have strong periodic behaviour that may be related to flu epidemics and to other long-term variations, it is necessary to remove this source of confounding before assessing the effect of PM10 on health. The usual strategy is to include in the models a nonparametric trend by making use of smoothing splines, giving rise to semiparametric models. Because of the close relationship between DGLMs and semiparametric GLMs (Fahrmeir and Tutz, 1994), we decided to control the long-term variations by introducing a stochastic trend modelled by a random walk of order 2, which allows us to compare our results readily with those already published.

Moreover, as we assumed that the counts reflect an underlying tendency of the severe air pollution events, combined with adverse meteorological conditions, to cause non-accidental death, we included in the \( x_t \) component vector those covariates which measure exposure to pollution and meteorological conditions, i.e. minimum temperature (Tmin), maximum temperature (Tmax) and the average daily specific humidity (Hum). We chose to include the pollutant’s concentration at lag 1, as Smith’s analyses (Smith et al., 2000) showed a spurious association between the response and the pollutant’s concentration at lag 0. Therefore, we initially set \( x_t = (T_{max}, T_{min}, Hum, PM10_{t-1})' \). We adopted the simplest formulation for describing the dynamic of the coefficients, i.e. a random walk of order 1. Intermediate analyses highlighted the non-significant effect of Tmax, so this variable was removed from the model. Moreover, as confidence bounds about the trajectory of the effect of Tmin, did not show a variation in time, although they indicated significance of the minimum temperature, in the subsequent analyses the parameter of this covariate was treated as a ordinary fixed effect.
As the cross-validation criterion that is used to estimate the hyperparameter $\lambda$ may have local minima, the results could be sensitive to the choice of the starting values for the optimization algorithm. Therefore, we decided to perform several pilot runs and to choose the initial values for the hyperparameters, i.e. variances of the error terms in the random walks, leading to the minimum observed value for the cross-validation criterion. The parameters were set as follows: $\sigma_\xi^2 = 7.1 \times 10^{-7}$ and $\Sigma_\xi = \text{diag}(4.6 \times 10^{-7}, 3.2 \times 10^{-11}, 1.8 \times 10^{-8})$. Moreover, the two initial values for the process $\{\omega_t\}$ were set equal to the estimated intercept of a GLM on the same regressor term.

6. Results

All analyses were carried out using R functions and C routines written by the authors. Fig. 2 shows the time series of the observed counts along with the fitted values, whereas Fig. 3 shows the estimated stochastic trend. For comparison, a cubic smoothing spline is also presented. The estimated spline was obtained from a semiparametric model including a smoothing spline for the trend component and the same explanatory variables as our final model, i.e. from the model

$$\text{log}(\mu_t) = \alpha_1 + s(t, k) + \alpha_2 \text{Tmin}_t + \alpha_3 \text{Hum}_t + \alpha_4 \text{PM10}_{t-1}$$

(7)

where $s(t, k)$ denotes the smoothing spline with $k$ degrees of freedom. Note that the use of the generalized cross-validation criterion to select the degree of smoothness of the spline underestimated the smoothing parameter (overestimated $k$), which might be explained by the autocorrelation of the data (Wang, 1998). A better fit for the semiparametric model was obtained by
employing the Akaike information criterion which led to a choice of 15 degrees of freedom for the spline. Although the criterion that was chosen to select the penalty degree is not the same in the two modelling strategies, both estimates are very close, as shown in Fig. 3.

Fig. 4 shows the estimated coefficients for the relevant variables along with pointwise 95% confidence intervals in the DGLM setting. Inspection of Fig. 4 highlights the time-varying effect of humidity, which appears to be significant in the central period of time. The roughness of the estimate and of the relative confidence band indicate a small carry-over effect of the variable, showing that the effect at time $t$ depends on previous values of the variable. The effect of the pollutant is also time varying and it is significant around the middle of the study period. Interestingly, the time intervals of significance for humidity and PM10 partly overlap, which seems to confirm Smith’s intuition about the importance that is played by humidity in the association between PM10 and health. In several runs using different starting values for the optimization algorithm, the significance and dynamic patterns of the estimates never varied greatly. Therefore, we are reasonably confident about the stability of our results.

For comparison, we report here the estimates $\hat{\alpha}_i, i = 1, \ldots, 4$, along with corresponding $t$-values, obtained for the coefficients of semiparametric model (7): $\hat{\alpha}_1 = 2.659 (86.86)$; $\hat{\alpha}_2 = -0.008 (-3.14)$; $\hat{\alpha}_3 = 0.008 (1.66)$; $\hat{\alpha}_4 = 0.001 (2.65)$. The conclusion from this analysis is that both minimum temperature and PM10 have a significant effect, whereas humidity is not significant. The size of the effects estimated in model (7) are comparable with those estimated in the DGLM. In this respect, the results obtained from the two approaches are consistent. However, evidence about the statistical significance of the effects partly disagrees. As already noted by Smith et al. (2000), standard models like model (7) are highly sensitive to the definition
of the exposure measure. For example, significance of the pollutant might easily disappear by making use of distributed lags for PM10 to take into account the carry-over effect.

To assess the overall adequacy of our DGLM, a graphical assessment of normality and independence of the randomized residuals was conducted. An examination of Fig. 5 reveals no evidence of departure from normality. Moreover, no evidence of autocorrelation is shown by the autocorrelation function of Fig. 6, as confirmed also by the non-significance of the Ljung–Box test (LB = 33.67 for lag 30) and of a Durbin–Watson test (DW = 2.02).

7. Discussion

We have adopted a dynamic generalized linear modelling approach to deal with time series regression of epidemiological time series. Our main objective was to explore the potential of this modelling framework to accommodate complex dependence structures while keeping transparency of the models and interpretability of the results. The example using the Birmingham data demonstrates significant potential for this modelling approach and illustrates how just a few variables can provide a useful insight into the structure of the data without compromising the fit.

We believe that this framework provides a valuable addition to the existing methodology for gaining insight into the temporal features of the data. In the example, we have detected, as in other studies, a positive effect of particular matter on mortality, but we have found that the significance of this effect might change in time. This might explain some of the controversy high-
Fig. 5. Diagnostic plots on transformed $P$-scores: (a) scores versus time and (b) normal quantile–quantile plot.

Fig. 6. (a) Estimated autocorrelation and (b) partial autocorrelation function of the transformed $P$-scores.
lighted by some reanalyses of these data based on different strategies for controlling temporal and confounding effects.

With respect to the pollutant’s effect, it might be argued that there is no physical reason for this coefficient to show dynamic behaviour, and that the dynamic behaviour is more an indication of a ‘lack of fit’, rather than a picture of reality. Even if this was the case, most of the covariates which would be needed to motivate a fixed effect for the pollutant are typically not available and some of them are probably not even detectable, so a dynamic framework naturally allows a flexible structure to incorporate the effects of unobserved factors on the response. To provide an example, consider for example the ‘harvesting hypothesis’, i.e. the hypothesis according to which only extremely frail individuals are affected by air pollution (Zeger et al., 1999). This means that increased mortality is associated with higher pollution levels only in the restricted group of people for whom life expectancy is short in the absence of pollution. An appropriate regression model may thus have a time-varying offset representing the size of the pool of the at-risk frail individuals, which is typically not available. However, imagine that two successive days with high pollution occur, sharing also the same values for the remaining covariates. If the harvesting hypothesis were true then the predicted number of deaths on the second day should be smaller than that on the first day, because the previous day’s high mortality would deplete the pool. This implies that we would need to have a model which allows the prediction of different outcomes for two subsequent days sharing the same covariates. In standard settings, this is not possible. DGLMs, which allow changes in the pollutant’s coefficient between time $t$ and time $t+1$, provide a useful tool to explore the harvesting hypothesis. For an alternative approach, see for example Schwartz (2001).

Computationally, our approach is more demanding than approaches that are typically adopted in similar applications. This disadvantage is offset by the conceptual appeal of controlling long-term and short-term dynamics and of avoiding difficult and not always transparent ad hoc selections of the time lag structure for the covariates and by the possibility of deriving residuals with clear properties, enabling greater confidence in model checking.

**Acknowledgements**

This work was supported by Ministero dell’Università e della Ricerca Scientifica e Tecnologica (Italy) under grant 9813200379, ‘Statistical methods for monitoring the environment and the health–environment interactions’, and by Ministero dell’Istruzione, dell’Università e della Ricerca (Italy) under grant MM3208412, ‘Statistics in environmental risk assessment’. We thank Adelchi Azzalini for helpful discussions on a preliminary version of this paper. We also thank the referees, the Associate Editor and the Joint Editor for comments which have led to substantial improvements of the paper.

**Appendix A**

In this section we shall describe the WKFS proposed by Fahrmeir and Wagenpfeil (1997).

Assume first that the hyperparameter $\lambda$ is known. If models (3) and (4) are sufficiently well behaved, the conditional mode $\hat{\alpha}_T$ is the solution of the vector equation

$$\frac{\partial \text{PL}(\alpha_T^*)}{\partial \alpha_T^*} = 0,$$

so that $\hat{\alpha}_T^*$ is a solution to the equations
for $t = 1, \ldots, T$, where $d_t = 1$ for $t = 1, \ldots, T - 1$ and $d_T = 0$. Following Fahrmeir and Wagenpfeil (1997), we derive an approximate solution of equations (8). Suppose that $\theta_t = \zeta_t' \alpha_t$, as in log-linear Poisson models, and consider
\[
\frac{\partial \log \left\{ p(y_i | \alpha_t) \right\}}{\partial \alpha_t} = z_i' \{ y_i - \mu_t(\alpha_t) \}.
\]
Furthermore, suppose that $\hat{\alpha}$ is a trial value of $\alpha_t^*$. Letting $\tilde{\theta}_t = \zeta_t' \hat{\alpha}$, and expanding expression (9) around $\tilde{\theta}_t$ yields
\[
\frac{\partial \log \left\{ p(y_i | \alpha_t) \right\}}{\partial \alpha_t} \approx z_i' \{ y_i - \mu_t(\hat{\alpha}_t) - \Sigma_t^{-1}(\hat{\alpha}_t) (\zeta_t' \alpha_t - \zeta_t' \hat{\alpha}_t) \}.
\]
By setting
\[
\begin{align*}
\tilde{H}_t &= \Sigma_t^{-1}(\hat{\alpha}_t), \\
\tilde{y}_t &= \Sigma_t^{-1}(\hat{\alpha}_t) \{ y_i - \mu_t(\hat{\alpha}_t) \} + z_t' \hat{\alpha}_t,
\end{align*}
\]
we can obtain the following linearized form for equations (8):
\[
-Q_t^{-1}(\alpha_t - F_t \alpha_{t-1}) + d_t F_{t+1} Q_{t+1}^{-1}(\alpha_{t+1} - F_{t+1} \alpha_t) + z_t H_t^{-1}(\tilde{y}_t - \zeta_{t} \alpha_{t}) = 0,
\]
which can be solved for $\alpha_t^*$ by the Kalman filter and smoother.

Let $\alpha = (\alpha_0', \alpha_1', \ldots, \alpha_T')'$ be a starting value. A complete iteration of the WKFS is as follows.

(a) Initialization: $\alpha_{00} = a_0$; $V_{00} = Q_0$.
(b) Filtering: for $t = 1, \ldots, T$,
\[
\begin{align*}
\alpha_{t|t-1} &= F_t \alpha_{t-1|t-1}, \\
V_{t|t-1} &= F_t V_{t-1|t-1} F_t' + Q_t, \\
K_t &= V_{t|t-1} \zeta_t' \{ \zeta_t' V_{t|t-1} \zeta_t + \Sigma_t^{-1}(\hat{\alpha}_t) \}^{-1}, \\
\tilde{y}_t &= \Sigma_t^{-1}(\hat{\alpha}_t) \{ y_i - \mu_t(\hat{\alpha}_t) \} + z_t' \hat{\alpha}_t, \\
\alpha_{t|t} &= \alpha_{t|t-1} + K_t (\tilde{y}_t - \zeta_t \alpha_{t|t-1}), \\
V_{t|t} &= V_{t|t-1} - K_t \zeta_t' V_{t|t-1}.
\end{align*}
\]
(c) Smoothing: for $t = T, \ldots, 1$,
\[
\begin{align*}
B_t &= V_{t-1|t-1} F_t' V_{t|t-1}^{-1}, \\
\alpha_{t-1|T} &= \alpha_{t-1|t-1} + B_t (\alpha_{t|T} - \alpha_{t-1|t-1}), \\
V_{t-1|T} &= V_{t-1|t-1} + B_t (V_{t|T} - V_{t-1|t-1}) B_t'.
\end{align*}
\]
For $\hat{\alpha} = (a_0, \alpha_1', \ldots, \alpha_T')'$, the WKFS specializes to the generalized extended Kalman filter and smoother (Fahrmeir, 1992). As we want to solve equations (8) iteratively, we must iterate the WKFS in the following way:

(a) compute $\hat{\alpha}^0$ with the generalized extended Kalman filter and smoother;
(b) repeat set $\hat{\alpha} = \hat{\alpha}^k$, compute $\hat{\alpha}^{k+1}$ with the WKFS and set $k = k + 1$ until some stopping criterion applies.

Estimation of the hyperparameter $\lambda$ can be accomplished by employing the cross-validation criterion
\[
\text{GCV}(\lambda) = \frac{1}{T} \sum_{t=1}^{T} \frac{\{ y_t - \mu_t(\hat{\alpha}_t) \}^2 / \mu_t(\hat{\alpha}_t)}{\{1 - \text{tr}(S_\lambda)/T\}^2},
\]
where $\hat{\alpha}_T^* = (\hat{\alpha}_1, \ldots, \hat{\alpha}_T)'$ is the solution of equations (8) for a fixed $\lambda$ and
\[
\text{tr}(S_\lambda) = \sum_{t=1}^{T} \mu_t(\hat{\alpha}_t) z_t V_{t|T} z_t'.
\]
\( \mu_t(\alpha_t) \) and \( V_{t|T} \) depend on \( \lambda \) and can be obtained from the WKFS without any additional computational effort. The quantities \( V_{t|T} \) computed by the WKFS are the diagonal blocks of the inverse of the matrix 

\[ -E \{ \delta^2 PL(\alpha_t^0) / \delta \alpha_t^0 \delta \alpha_t^0 \} \]

Therefore, \( V_{t|T} \) can be interpreted as the error covariance matrix and used to build approximate pointwise confidence intervals for \( \alpha_t, t = 1, \ldots, T \).

References


